REPLICATION FILES

Alexander Gelber, Timothy J. Moore, Zhuan Pei and Alexander Strand.

"Disability Insurance Income Saves Lives," Journal of Political Economy

Production Date: February 1, 2023

Programs used: SAS9.4 and Stata17

==========

OVERVIEW

This replication package primarily constructs the analysis files from administrative data sources from the Social Security Administration (SSA). The codes provided generate all tables and figures in the paper and in the online appendices.

A replicator interested to reproduce the analysis in the manuscript would need to apply for the data from SSA. We are happy to assist interested researchers to apply. The most relevant authors for providing this assistance can be reached at moore839@purdue.edu and alexander.strand@ssa.gov.

There are a small number of additional statistics reported in the text using data from the Survey of Income and Program Participation (SIPP) and Chetty et al. (2016). This replication package also provides the code for these, and the publicly available files when they are ones used.

Details about the non-SSA data and code are provided at the end of the document in their own sections, as they are only used to provide specific information in footnotes (not for any of the figures or tables)>

==========

CODEBOOKS (\1_documentation)

We use data from the Social Security Disability Analysis File for 2010 (formerly the Ticket Research File, or TRF10), Master Earnings File, Numident File and Continuous Work History Sample. In the "1_documentation" folder, we provide codebooks for these data. Our variable names match this documentation. Note that we use both the Demographic File and the Annual Files (from 1997 to 2010).

We provide the following codebooks in PDF format:

- TRF10_ConstructionGuide_Vol1.pdf, TRF10_ConstructionGuide_Vol2.pdf
- TRF10_UsersGuideVol1.pdf, TRF10_UsersGuideVol2.pdf
- TRF10DataDictionaryVol1.pdf, TRF10DataDictionaryVol2.pdf
- Continuous Work History Sample.pdf
- Masters Earning File.pdf
- Numident File.pdf

Note that more background information on these data and the relevant variables are available at the home page for the Disability Analysis File:
https://www.ssa.gov/disabilityresearch/daf.html

==========

DATA (\2_data)

In addition to the SSA data (once obtained), the “2_data” folder is the home for cleaned data sets used in the analysis. These are listed as outputs below for the relevant code. This folder must be created.

==========

CODE (\3_code)

In order to prepare the data, please use the files in the "2_data" folder or designate a file path to the source data. This will need to be done for the SSA data.

Once that is done, please run the files in the “3_code” folder. The order of the files is as follows:

- 1_cleaning_data.sas reads the Disability Analysis Files, Master Earnings File and Numident File and calculates the timing of filing, the start of payments (including dependent benefits), mortality in relation to the start of benefits, earnings in relation to the start of benefits, and available covariates. It also cleans data in a similar way for non-beneficiaries using the Continuous Work History Sample.

  SAS was used to deal with large files at this stage of the process – the rest of the code is in Stata. The following cleaned Stata data sets are produced:

  \2_data\di_file.dta
  \2_data\nondi_file.dta

- 2_creation of SSA_samples.do further cleans the data and creates the analytic sample files:

  \2_data\lower_main_sample.dta
  \2_data\lower_nonbenef_sample.dta
  \2_data\lower_ssi_sample.dta

  \2_data\fm_main_sample.dta
  \2_data\fm_nonbenef_sample.dta
  \2_data\fm_nondepend_sample.dta
  \2_data\fm_ssi_sample.dta

  \2_data\upper_main_sample.dta
  \2_data\upper_nonbenef_sample.dta
  \2_data\upper_ssi_sample.dta

- 3_figures.do creates all of the figures in the same order in the paper using the SSA microdata, which is all of the figures except those constructed directly from SSA reports to describe the policy variation (Figure 1, Figure A1, Figure A2). The resulting figures are outputted to \4_output\figures and the log file is outputted to \4_output\log.

- 4_tables.do creates all of the tables in the same order as in paper using the SSA microdata, which is all of the tables except one constructed directly from SSA reports to describe the policy variation (Table A1). The resulting tables are outputted to \4_output\tables and the log file is outputted to \4_output\log.

==========

OUTPUT (\4_output)

This is the location for all figures and tables in the paper produced using microdata. The Stata do file figures.do produces the figures into \4_output\figures. The Stata do file tables.do produces the tables into \4_output\tables. Each do file produce a single log file, which it outputted to \4_output\log.

==========

SIPP SYNTHETIC BETA (\5_synthetic_SIPP)

This is the code that produces the comparisons across the bend points discussed in footnote 18.

The SIPP Synthetic Beta (SSB) is a Census Bureau product that integrates person-level micro-data from the SIPP household survey with administrative tax and benefit data. Details on the structure, variables and data access are provided here:

https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html

Once access is granted, the code in this folder should be supplied to the Census staff to reproduce our findings. Both sets of code should be submitted to them. Running 1_SIPP_synthetic_master.do will use 2_SIPP_synthetic_create.do and produce the desire output.

==========

SIPP PUBLIC (\6_public_SIPP)

This is the code that produces the scaling discussed in footnote 19. The SIPP Synthetic Beta does not measure family income (and either does the SSA microdata), so to scale by family income we use the SIPP public-use data for this purpose. The do files extract and calculate descriptive statistics from the SIPP:

- 1_download_and_prep_SIPP.do creates and cleans the SIPP dataset, 1996_to2008_footnote19.dta, which we include in the folder
- 2_sipp_family_income_to_individual_income_ratios.do creates the highlighted ratios to scale our effects in that footnote.

The information produced is provided in footnote19_information.xlsx.

==========

CHETTY ET AL 2016 (\7_chetty_etal)

This is the code that produces the correlations from Chetty et al. (2016) that are discussed in footnote 22. We download the data from: https://opportunityinsights.org/paper/lifeexpectancy/, and include it (health_ineq_online_table_15.csv). Then use chetty-etal-analysis.do to generate the estimates of the relationship between household income and mortality in their data, which are converted to percentage estimates by using the mortality rate for the lowest percentile of taxpayers in their household income.

==========